



Bioinformatics: An Emerging Means of Assessing Environmental Health

In a recent editorial by Lucier (*EHP* 101:366) concerning issues of risk assessment, inadequacies of contemporary basic and applied research were discussed in relation to the high cost to government and industry. A plea for better communication between scientists and regulators was deemed necessary, along with a national policy on both scientific and regulatory aspects of risk assessment. Writing as a basic scientist with a broad background in biology, I know that methods are now available for altering the course of research impinging on risk assessment. These methods, which employ computer-driven "bioinformatics" coupled with the ability to rapidly obtain partial sequences of expressed genes (about 1000 expressed genes/day), portend revolutionary changes in both scientific research and medical practices and how data can be collated for useful regulatory purposes.

The underlying biological basis for using these new procedures is the fact that all living organisms strive to react to their environment using every adaptation mechanism in their repertoire. The culminated expression of formerly latent genes with the concomitant appearance of new cellular architecture and different sets of sensory/regulatory mechanisms modulate the effects originating from the environment. The plasticity of cellular responses is recorded by the reversibility of gene expression subsequent to withdrawal of the environmental signals impacting the organism. Alternatively, toxic environmental signals may lead to defects in genes, ending with expressed gene patterns that are not reversible. Hence, both reversible and irreversible changes in total expressed gene patterns can be used for quantitative and qualitative assessment of the potential hazards of environmental chemicals to a particular cell, organ, or animal.

Yeast provides an excellent model organism for demonstrating the usefulness of bioinformatics for rapidly evaluating the effects of the environment on the expressed gene patterns and consequent metabolic and structural responses. For example, when grown on a rich culture medium replete with all the necessary ingredients for rapid growth, the expressed gene pattern and the equivalence in expressed proteins (using appropriate databases and informatic programs) display the expected richness in metabolic enzymes. When

grown on a low-carbon source with minimal nutritional ingredients, these cells yield a markedly different pattern of expressed genes that produce proteins not expressed in the rich culture medium. Some of the proteins are typically stress-induced proteins in all animal cells (for example, heat-shock proteins). Hence, two-distinct "gene-libraries" can be obtained in a brief period that serve as useful markers for the metabolic and structural patterns given under two extreme conditions.

The Institute of Genomic Research (a basic research institute that pioneered many of the techniques employed in bioinformatics), among other similarly directed institutions, intends to sequence at least 90% of the expressed human genes by the fall of 1994. With already obtained extensive cDNA libraries from many types of human cells and tissues, including malignant cells, it takes little imagination to foresee that it will soon be possible to distinguish expressed gene patterns between distinct human cell types and to determine, under appropriate culture conditions with specific cells, whether a potentially hazardous compound produces a reversible or irreversible pattern of total expressed genes from "normal." Assuming that irreversibility of gene patterns is an indication of hazard to the cell and possibly to the human body, dose-response relationships under appropriate culture conditions with one or more cell types can be accurately determined. The important difference from other toxicological approaches is that the method involves the total expressed gene pattern. Specific genes or regulatory pathways may be determined by the techniques but are not immediately or necessarily important to the final assessment.

With broadened databases, the ability to correlate protein structure and function, and with accurate gene mapping of the affected genes, the gateway will be opened in the near future to new vistas for research at both basic and applied levels. Most importantly, given the relatively brief time that it has taken to achieve the present status of bioinformatics coupled with the enormous compilation of information already gathered on biological systems encompassing microorganisms, plants, and animals, it is essential that government, industry, and academic institutions prepare for what undoubtedly will be a revolution in thought and practice.

Martin Rodbell

National Institute of Environmental Health Sciences